融合 IMU 去除运动模糊的改进光流匹配算法 *

栾小珍 [a]，魏国亮 [b†]，蔡　洁 [a]
(上海理工大学 a. 理学院; b. 管理学院, 上海 200093)

摘　要: 为进一步提高视觉 SLAM 中的光流匹配精度和速度,提出一种融合惯性测量单元(Inertial Measurement Unit, IMU)去除运动模糊的改进光流匹配算法。该算法首先利用 IMU 运动信息计算的点扩散函数去除运动模糊，提高特征点匹配率；其次在 LK(Lucas-Kanade)光流的基础上，引入梯度误差，并使用图像梯度 L1 范数作为正则项模拟稀疏噪声，构建代价函数；然后利用 IMU 预测特征点位置作为该算法初始值，并加入 BB(Barzilar-Borwein)步长改进原有的高斯牛顿算法，提高计算速度。实验表明，通过两帧之间比较，该算法的效率和精度均优于 LK 光流法；然后将该算法集成到 VINS-Mono 框架，在数据集 EuRoC 上，结果显示该算法提高了原有框架的定位精度和鲁棒性。
关键词： 光流法；前端视觉里程计；运动模糊；多传感器融合
中图分类号： TP242　　doi: 10.19734/j.issn.1001-3695.2022.02.0075

Feature tracking algorithm for removing motion blur by improving optical flow incorporating IMU

Luan Xiaozhen[a], Wei Guoliang[b†], Cai Jie[a]
(*a. College of Science*, *b. Business School*, *University of Shanghai for Science & Technology*, *Shanghai* 200093, *China*)

Abstract: In order to improve the accuracy and efficiency of the feature point matching, this paper proposed a novel feature point matching algorithm in terms of vision and inertial measurement unit (IMU) fusion. Firstly, the algorithm calculated the point diffusion function using the motion information of IMU to remove motion blur, and improved the feature point matching rate. Secondly, based on LK (Lucas-Kanade) optical flow method, this paper introduced gradient error and uses L1 parametric to simulate sparse noise. Furthermore, the feature point position by using IMU is the initial value of the algorithm, and then this paper used BB (Barzilar-Borwein) step to improve the efficiency of the algorithm. Finally, the comparison experiments show that the efficiency and accuracy of the algorithm are better than the LK optical flow method. Especially, the algorithm improves the localization accuracy and robustness of the VINS-Mono framework on the dataset EuRoC.
Key words: optical flow; visual odometry; motion blur; multi-sensor fusion

0　引言

近年来，视觉同时定位与建图(Visual Simultaneous Localization And Mapping，V-SLAM)[1]在机器人领域有着重要的作用，如物流配送、虚拟现实(Virtual Reality，VR)[2]、智能家居等。虽然单目相机成本较低、使用方便，但是仅仅依靠相机，难以估计环境的尺度。因此在 V-SLAM 中，为了解决尺度不可观问题，众多学者更倾向于多传感器融合技术。由于惯性测量单元(Inertial Measurement Unit，IMU)能够在隔绝外部环境的影响下测量本体的角速度和加速度，从而使单目视觉的公制尺度可观测，并提供鲁棒和准确的帧间运动估计，因此视觉和 IMU 融合的 SLAM 技术(VI-SLAM)异军突起[3-5]。VI-SLAM 框架主要分为前端里程计、后端优化、回环检测和建图。前端里程计作为首要步骤，在整个系统中显得尤其重要。在前端里程计中，不仅需要进行 IMU 预积分，而且需要对相机传入的图片进行特征点提取和匹配，进而根据匹配信息计算相邻两帧之间机器人的运动，并构建特征点的局部 3D 位置[6]。一旦发生误匹配[7]，会导致估计不准确，因此，特征点提取与匹配也是前端里程计中的重要一环。

特征点提取与匹配主要分为特征点法[8,9]、光流法[10]和直接法[11]。特征点法在计算关键点、描述子，以及匹配描述子时计算量过大，而光流法和直接法弥补了以上缺陷，光流法只计算关键点，不计算描述子，通过两帧之间的光度不变假设，直接最小化光度误差并进行迭代匹配，本文重点研究光流法。

在光流法中目前主要存在以下两个问题，一方面，若物体运动过快，则在相机曝光期间，景物和摄像机之间存在相对运动，会导致像素重叠，出现图像模糊现象，称为运动模糊。而光流法对光度稳定和图像的连续有较强依赖，当不满足以上假设的时候，容易找不到跟踪点，更糟糕的情况是追踪到错误的特征点。另一方面，在光流法中使用非线性优化方法进行最小化光度误差迭代匹配的过程中，首先，非线性算法的最优值对初值有很强的依赖性，如果没有初值则需要进行全局搜索，计算量较大，且容易造成误匹配；其次，传统高斯牛顿方法的步长是固定步长或者直接由线搜索算法给出，如果步长取的太大，一开始会很快向最优解逼近，在收敛的最后阶段，容易跳过最优解；而步长取的太小，则导致收敛速度变慢。因此，去除运动模糊和选取非线性优化算法的初始值、步长是本文研究的两个问题。

基于以上问题，本文提出了一种融合 IMU 去除运动模糊的光流匹配算法。该算法通过融合 IMU 运动信息，计算点扩散函数去除运动模糊，之后再次融合 IMU 信息预测特征点位置，为光流跟踪提供初值，进而提高特征点提取与匹配的性能。最后，将该算法集成到了完整 VI-SLAM 框架中。

在 VI-SLAM 框架特征点提取与检测方面，Tong Qin 等提出的 VINS-Mono[12]前端使用光流法提取关键点，并使用多层金字塔 LK(Lucas-Kanade)进行跟踪，虽然整体框架是基于 IMU 和视觉传感器融合，但在关键点提取和跟踪中，并未将 IMU 信息与视觉信息融合。Qi Guan 等提出的 IFPT 特征点匹

收稿日期： 2022-02-23；**修回日期：** 2022-04-13　**基金项目：** 上海市“科技创新行动计划”国内科技合作项目(20015801100)
作者简介： 栾小珍(1997-)，女，山东聊城人，硕士研究生，主要研究方向为视觉 SLAM、计算机视觉；魏国亮(1973-)，男(通信作者)，河南新乡人，教授，博导，博士，主要研究方向为非线性系统优化、机器视觉、随机系统(guoliang.wei@usst.edu.cn)；蔡洁(1996-)，女，江苏徐州人，博士研究生，主要研究方向为视觉 SLAM.

配算法将特征点法与 IMU 进行融合[13]，通过 IMU 的位置和旋转方向预测下一帧关键点的位置，并以其为圆心，进行描述子局部搜索与匹配。OKVIS(Open Keyframe-based Visual-Inertial SLAM)[14]的 3D-2D 的匹配中，利用 IMU 预测的位姿筛选当前帧可见的特征点，去除掉当前帧已经追踪不到的特征点，节省暴力匹配时间。OKVIS 和 IFPT 共同点是将特征点法与 IMU 融合，即基于预积分理论进行预测，为待匹配的每个特征点提供新图像中的可能位置，进而完成前端特征点提取与匹配。不同点是 IFPT 得到可能位置之后进行局部搜索，而 OKVIS 筛选去除当前帧已经看不到的点，然后进行全局暴力匹配。

本文提出的算法和之前相关工作有所区别，首先从算法本身比较：a) 之前的工作[15]直接设置平移信息 $t=0$，而提出的算法充分利用了 IMU 的信息，加入了加速度计得到的 t，实验结果表明融合平移信息后，IMU 预测的特征点位置更加准确；b) 之前的工作未将速度信息充分利用，而提出的算法利用 IMU 预积分得到的运动信息，计算点扩散函数，加入了运动模糊复原步骤，减少运动模糊导致匹配特征点失败现象；c) 在非线性优化方法中，加入 BB(Barzilar-Borwein)可变步长进行搜索，使步长大小可以自适应损失函数。其次，从框架上比较，之前的工作[15]把偏置当作状态向量的一部分去优化，并且在初始化前，偏置求解不准确的情况下，仍然预测相机位姿，导致视觉里程计的误差一直传递下去，更糟糕的是若系统中没有闭环检测，很难消除视觉里程计产生的累计误差。本文提出的算法利用 IMU 预测的时机是在系统初始化后，此时偏置信息已经求出，减小了累计误差。

1　系统框架

该算法在 VINS-Mono 框架前端算法基础上进行研究。原 VINS-Mono 框架前端算法使用 ShiTomasi[16,17]算法检测角点，LK 光流法跟踪匹配。在运动幅度较大，光照变化剧烈情况下，匹配误差较大。

为解决以上问题，提高定位精度，本文对 VINS-Mono 前端算法进行改进。具体实施方案如下：首先在初始化前，仍通过传统光流法进行特征点提取，初始化完成之后，将得到的 IMU 加速度计和角速度计偏置代入到预积分中，融合预积分预测得到的速度、位置、旋转方向信息，发送到前端，判断速度是否大于阈值，如果大于阈值说明运动过快，则根据运动信息进一步对图像进行非盲复原[18,19]。如果上一步完成或者速度小于阈值，说明运动幅度较小，为满足实时性，则不进行图像复原操作。直接将 IMU 预测得到的位置、旋转方向与视觉信息融合，即将恢复之后的点投影到下一帧图像上，提供给光流追踪中高斯牛顿算法一个初始值，确定光流追踪匹配的局部搜索位置。该算法集成到框架上，其流程如图 1 所示。

2　预积分模型预测特征点位置

本节首先对常用到的一些符号进行定义。$\hat{(\cdot)}$ 表示传感器测量值，R 表示 3×3 的旋转矩阵，q 表示旋转的四元数表示，t 表示 3 维平移向量，$T=[R\,|\,t]$ 表示姿态的转换矩阵，包括旋转矩阵和平移向量。$(\cdot)_I$ 为 IMU 坐标系，$(\cdot)_C$ 为相机坐标系，特别地，令相机的第一帧为参考坐标系，即旋转为单位矩阵 $R_{C_0}=E$，平移向量为零向量 $t_{C_0}=0$。假设经过初始化[20,21]之后，从相机坐标系到 IMU 坐标系的旋转外参 R_{bc}，平移外参 t_{bc}，IMU 陀螺仪的偏置 b^g，加速度计偏置 b^a，以及在参考帧坐标系下的重力矢量 g_{C_0} 是已知的。

2.1　基于预积分求解机器人的位置、速度、旋转方向

只考虑陀螺仪和加速度计的偏置白噪声以及偏置的随机游走噪声，则 IMU 模型为[22]：

$$\begin{aligned}\hat{w}_{C_0I_j}&=w_{C_0I_j}+b^g+n_j^g\\ \hat{a}_{C_0I_j}&=q_{C_0I_j}^T\,(a_{C_0I_j}+g_{C_0})+b^a+n_j^a\end{aligned}\tag{1}$$

其中：$w_{C_0I_j}$ 为陀螺仪在 IMU 坐标系下真实的角速度值，$\hat{w}_{C_0I_j}$ 为陀螺仪在 IMU 坐标系下测量得到的角速度值，$a_{C_0I_j}$ 为加速度计在 IMU 坐标系下真实的加速度值，$\hat{a}_{C_0I_j}$ 为加速度计在 IMU 坐标系下测量得到的加速度值。b^g，b^a 的导数服从高斯分布：

$$\dot{b}^g\sim\mathbb{N}(0,\sigma_{b^g}),\ \dot{b}^g\sim\mathbb{N}(0,\sigma_{b^a});$$

其中：n_j^g，n_j^a 分别为陀螺仪、加速度计偏置的随机游走噪声，服从高斯分布：

$$n_j^g\sim\mathbb{N}(0,\ \sigma_g),\ n_j^a\sim\mathbb{N}(0,\ \sigma_a);$$

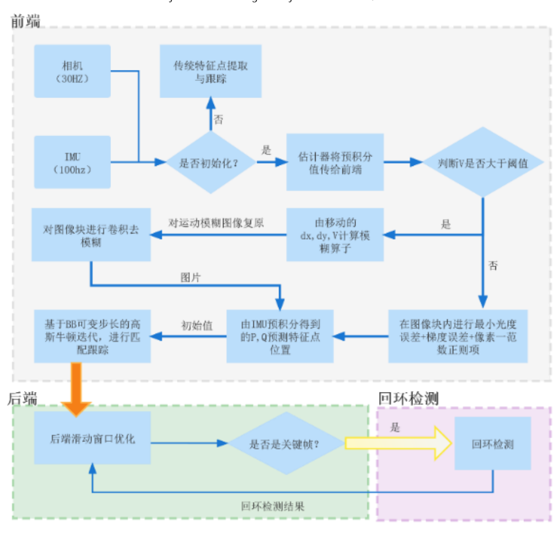


图 1　系统框架

Fig. 1　System framework

在相机坐标系下，分别将 j 时刻机器人位置、速度、旋转方向对时间分别求导得：

$$\begin{aligned}\dot{p}_{C_0I_j}&=v_{C_0I_j}\\ \dot{v}_{C_0I_j}&=a_{C_0I_j}\\ \dot{q}_{C_0I_j}&=q_{C_0I_j}\otimes\begin{bmatrix}0\\ \frac{1}{2}w_{C_0I_j}\end{bmatrix}\end{aligned}\tag{2}$$

其中，$p_{C_0I_j}$ 为从相机第一帧坐标系到 IMU 第 j 帧坐标系的平移向量；$v_{C_0I_j}$ 为相机坐标系下，j 时刻机器人速度，用三维向量表示；$q_{C_0I_j}$ 为从 IMU 第 i 帧坐标系到相机第一帧坐标系的旋转矩阵，用四元数表示。$\dot{p}_{C_0I_j}$、$\dot{v}_{C_0I_j}$、$\dot{q}_{C_0I_j}$ 表示将机器人的位置、速度、旋转方向分别对时间求导。

对式(2)进行积分得到在相机坐标系下，j 时刻机器人的位置、速度、旋转方向：

$$\begin{aligned}p_{C_0I_j}&=p_{C_0I_{j-1}}+v_{C_0I_j}\triangle t+\frac{1}{2}a\triangle t^2\\ v_{C_0I_j}&=v_{C_0I_{j-1}}+a\triangle t\\ q_{C_0I_j}&=q_{C_0I_{j-1}}\otimes\begin{bmatrix}1\\ \frac{1}{2}w\triangle t\end{bmatrix}\end{aligned}\tag{3}$$

其中 Δt 为 i 时刻与 j 时刻的时间间隔，采用中值法得到 j 时刻的加速度 a 和角速度 w：

$$\begin{aligned}a&=\frac{1}{2}\left[q_{C_0I_{j-1}}(a_{C_0I_{j-1}}-b^a)-g_{C_0}+q_{C_0I_j}(a_{C_0I_j}-b^a)-g_{C_0}\right]\\ w&=\frac{1}{2}\left[(w_{C_0I_{j-1}}-b^g)+(w_{C_0I_j}-b^g)\right]\end{aligned}\tag{4}$$

但是在优化位姿过程中，当前帧的速度和旋转矩阵等会发生改变，需要重新进行积分，计算量较大，故引入预积分[22,23]：

$$q_{C_0I_j}=q_{C_0I_k}\otimes q_{I_kI_j}\tag{5}$$

其中，$\otimes$ 表示四元数乘法[22]，将(5)代入(4)更新加速度：

$$a=q_{C_0I_k}\otimes\frac{1}{2}\left[q_{I_kI_{j-1}}(a_{C_0I_{j-1}}-b^a)+q_{I_kI_j}(a_{C_0I_j}-b^a)\right]-g_{C_0}$$

引入预积分之后运动学模型的离散形式：

$$
\begin{aligned}
p_{C_0I_j} &= p_{C_0I_{j-1}} + v_{C_0I_j}\Delta t - \frac{1}{2}g_{C_0}\Delta t^2 + q_{C_0I_k}\otimes\alpha_{I_kI_j}\\
v_{C_0I_j} &= v_{C_0I_{j-1}} - g_{C_0}\Delta t + q_{C_0I_k}\otimes\beta_{I_kI_j}\\
q_{C_0I_j} &= q_{C_0I_{j-1}}\otimes\gamma_{I_kI_j}
\end{aligned} \tag{6}
$$

其中：

$$
\begin{aligned}
\alpha_{I_kI_j} &= \frac{1}{2}\left[q_{I_kI_{j-1}}(a_{C_0I_{j-1}} - b^a) + q_{I_kI_j}(a_{C_0I_j} - b^a)\right]\Delta t^2\\
\beta_{I_kI_j} &= \frac{1}{2}\left[q_{I_kI_{j-1}}(a_{C_0I_{j-1}} - b^a) + q_{I_kI_j}(a_{C_0I_j} - b^a)\right]\Delta t\\
\gamma_{I_kI_j} &= \begin{bmatrix} 1 \\ \frac{1}{2}((w_{C_0I_{j-1}} - b^g) + (w_{C_0I_j} - b^g)) \end{bmatrix}
\end{aligned} \tag{7}
$$

为预积分量，计算式(6)得到了相机坐标系下机器人的位置、速度、旋转方向。

2.2 预测位姿以及特征点位置

结合 2.1 节得到的结果计算相机坐标系下 i 和 j 两帧之间的转换矩阵 $T_{C_jC_i}$：

$$T_{C_0C_j}T_{C_jC_i} = T_{C_0C_i} \tag{8}$$

其中

$$T_{C_0C_i} = T_{C_0I_i}T_{I_iC_i},\quad T_{C_0C_j} = T_{C_0I_j}T_{I_jC_j} \tag{9}$$

将式(9)代入式(8)化简分别得到 i 和 j 两帧之间旋转矩阵 $R_{C_jC_i}$ 和平移向量 $t_{C_jC_i}$：

$$
\begin{aligned}
R_{C_jC_i} &= R_{IC}^{-1}R_{C_0I_j}^{-1}R_{C_0I_i}R_{IC}\\
t_{C_jC_i} &= R_{IC}^{-1}[R_{C_0I_j}^{-1}(R_{C_0I_i}t_{IC} + t_{C_0I_i} - t_{C_0I_j}) - t_{IC}]
\end{aligned} \tag{10}
$$

其中，R_{IC}^{-1} 为从 IMU 坐标系到相机坐标系的旋转矩阵；${R_{C_0I_j}}^{-1}$ 为从相机第一帧坐标系到 IMU 第 j 帧坐标系的旋转矩阵；$R_{C_0I_i}$ 为从 IMU 第 i 帧坐标系到相机第一帧坐标系的旋转矩阵；R_{IC} 为从相机坐标系到 IMU 坐标系的旋转矩阵；t_{IC} 为从 IMU 坐标系到相机坐标系的平移向量；$t_{C_0I_i}$ 为从相机第一帧坐标系到 IMU 第 i 帧坐标系的平移向量。

将第 i 帧的特征点归一化坐标 p_i，映射到第 j 帧。得到

$$p_j = R_{C_jC_i}\dot{p}_i + t_{C_jC_i} \tag{11}$$

其中 $\dot{p}_i$ 为 p_i 的齐次坐标，大小为 3×1 的向量，p_j 为 IMU 预测得到的特征点在第 j 帧下的归一化坐标。

3 IMU 去除运动模糊光流特征点跟踪算法

运动模糊图像的复原模型可以描述为一个清晰图像经过退化函数作用之后，再加上一个噪声项。设一幅清晰图像 $f(x,y)$，经过退化函数卷积，则产生的退化图像 $g(x,y)$，具体为

$$g(x,y) = f(x,y)*h(x,y) + n(x,y) \tag{12}$$

其中 $h(x,y)$ 称为模糊算子或点扩散函数，$*$ 表示卷积，$n(x,y)$ 为加性噪声。

复原模型可以根据有无运动先验分为盲复原和非盲复原。由于通过 IMU 得到的加速度和角速度可以确定运动信息，因此研究重点放在非盲复原模型。

3.1 融合 IMU 先验信息的运动模糊非盲复原模型

当两幅图像的时间戳，即时间间隔 t_{ij} 已知时，通过(7)得到了机器人在两帧之间的速度信息，进而利用(11)计算了特征点在两帧的归一化坐标 (u_i, v_i)、(u_j, v_j)，于是得到机器人运动模糊方向的角度和长度：

$$
\begin{cases}
\theta = \arctan\left(\dfrac{v_j - v_i}{u_j - u_i}\right)\\
L = v\times t_{ij}
\end{cases} \tag{13}
$$

其中 v 表示机器人速度，t_{ij} 为时间间隔。通过运动模糊的方向 θ 和长度 L，则可确定运动模糊的点扩散函数宽度 b、高度 h：

$$
\begin{cases}
b = L\times\cos\theta\\
h = L\times\sin\theta
\end{cases} \tag{14}
$$

点扩散函数 $h(x,y)$ 由以下隐函数表示为

$$
h(x,y) = \begin{cases}
y = [\frac{b}{h}\times x] & x = 0,\ 1,\ \ldots, h-1\\
x = [\frac{h}{b}\times y] & y = 0,\ 1,\ \ldots, b-1
\end{cases} \tag{15}
$$

根据点扩散函数建立模糊复原模型 $f(x,y)$ [24]为

$$f(x,y) = \arg\min_{f(x,y)}\|g(x,y) - f(x,y)\otimes h(x,y)\|_2 \tag{16}$$

其中 $\|g(x,y) - f(x,y)\otimes h(x,y)\|_2$ 为图像保真项，保证清晰图像经过点扩散函数作用后，与模糊图像一致。式(16)可通过快速傅里叶变换求解[19]。

3.2 改进原有光流法优化求解匹配特征点

本文在原有 LK 光流假设两帧图像之间同一特征点光度不变的基础上，首先，加入梯度不变假设：

$$\nabla I_1(x,y) = \nabla I_2(x+\Delta x, \Delta y),$$

充分利用特征点与相邻像素的信息，保证正确的匹配对。其次，为了防止过拟合，使用第二帧图像梯度 L_1 范数模拟稀疏噪声，构建正则化项，最终模型建立如下：

$$
\begin{aligned}
(\Delta x, \Delta y) = \arg\min_{(\Delta x,\Delta y)}(&\|I_1(x,y) - I_2(x+\Delta x, \Delta y)\|_2 +\\
&\alpha\|\nabla I_1(x,y) - \nabla I_2(x+\Delta x,\Delta y)\|_2 + \beta\|\nabla I_2(x,y)\|_1)
\end{aligned} \tag{17}
$$

其中：$I_1(x,y)$ 表示参考帧在像素坐标 (x,y) 处图像的灰度值；$I_2(x+\Delta x,\Delta y)$ 表示当前帧在像素坐标 $(x+\Delta x,\Delta y)$ 处图像的灰度值；$\nabla I_1(x,y)$ 表示参考帧在像素坐标 (x,y) 处图像的梯度值；$\nabla I_2(x+\Delta x,\Delta y)$ 表示当前帧在像素坐标 $(x+\Delta x,\Delta y)$ 处图像的梯度值；α 为梯度权重系数，β 为稀疏噪声项权重系数。

在 2.2 节中，已经通过 IMU 预测得到第二帧特征点的位置，将预测位置作为高斯牛顿非线性优化算法初始值，并在原来固定步长基础上加入 BB 可变步长进行迭代。

将式(17)非线性函数 $F(X)$ 进行泰勒展开至二阶：

$$
\begin{aligned}
F(X) = &F(X_k) + \nabla F(X_k)(X - X_k) +\\
&\frac{1}{2!}(X - X_k)^T H_{X_k}(X - X_k)
\end{aligned} \tag{18}
$$

其中，X 为优化变量，X_k 为优化变量在第 k 步的取值，$F(X_k)$ 为非线性函数 $F(X)$ 在 X_k 处的取值，$\nabla F(X_k)$ 为非线性函数 $F(X)$ 在 X_k 处的梯度值，H_{X_k} 为 $F(X)$ 在 X_k 处的海塞矩阵。

对式(18)求导，令其为 0，得优化变量 $X = X_k - H_{x_k}^{-1}\nabla f(X_k)$，则优化方向 d_k 为

$$d_k = -H_{x_k}^{-1}\nabla F(X_k) \tag{19}$$

下面对步长进行求解，对于一般地二次函数优化问题，

$$F(X) = \frac{1}{2}X^TAX + b^TX \tag{20}$$

其中 A 表示二次函数的系数矩阵，b 为偏置。对式(20)求导，得一阶导函数 $F'(X) = AX + b$，则第 k 步迭代梯度为

$$\nabla F(X_k) = AX_k + b \tag{21}$$

将式(21)代入式(19)，加入常数高斯牛顿迭代步长 g_k，近似迭代方向：

$$d_k = -H_{x_k}^{-1}\nabla F(X_k) \approx -g_kI\nabla F(X_k) = -(g_k^{-1}I)^{-1}\nabla F(X_k) \tag{22}$$

其中 I 为单位矩阵，定义自变量的差 S_{k-1}，梯度差 Y_{k-1}：

$$S_{k-1} = X_k - X_{k-1}, Y_{k-1} = \nabla F(X_k) - \nabla F(X_{k-1}) \tag{23}$$

对式(21)在第 k 与第 $k-1$ 处做差：

$$\nabla F(X_k) - \nabla F(X_{k-1}) = AX_k - AX_{k-1} = AS_{k-1} \tag{24}$$

则得到拟牛顿方程：

$$AS_{k-1} = (g_k^{-1}I)S_{k-1} = Y_{k-1} \tag{25}$$

在式(25)中令 $\delta = g^{-1}$，则第 k 步步长：

$$g_k^{-1} = \arg\min_{g}\frac{1}{2}\|S_{k-1}\delta - Y_{k-1}\|_2 \tag{26}$$

求得式(26)的最小二乘解：

$$g_k = \frac{S_{k-1}^TS_{k-1}}{S_{k-1}^TY_{k-1}} \tag{27}$$

由式(27)可以看出，g_k 仅与第 k 步自变量状态信息和相邻两步函数的梯度有关，与其他变量无关。

4　实验结果与分析

本文所有实验均在同一平台进行，平台主要配置为 AMD Ryzen 5 3500U with Radeon Vega Gfx 四核笔记本处理器，内存为 12GB，基准频率为 2.10GHz，Ubuntu18.04 操作系统。

在本节，首先单独对算法本身的性能进行评估，将提出的算法与 VINS-Mono 框架中匹配算法 LK 光流法、ORB 特征点检测算法对比，测试在两帧之间匹配运行时间和匹配精度。其次，为了进一步验证提出的优化算法的准确性以及在系统框架中的稳定性，将该算法集成到 VINS-Mono 系统框架，与原 VINS-Mono、PL-VIO[25]、VINS-Fusion[26]单目相机加 IMU(Mono-IMU)、VINS-Fusion 双目相机加 IMU(Stereo-IMU)进行对比，评价定位精度。

实验选用的数据集是机器人微型飞行器(Micro Aerial Vehicle，MAV)上收集的视觉惯性数据集 EuRoC，飞行场景是苏黎世联邦理工学院 ETH 的一个机器工厂和两个普通房间场景。为了验证所提出算法在普通场景数据下的有效性，该算法使用了 3 个场景不同难度的 9 个序列：MH_02_easy，MH_03_medium，MH_04_difficult，V1_01_easy，V1_02_medium，V1_03_difficult，V2_01_easy，V2_02_medium，V2_03_difficult。

4.1　两帧之间算法运行时间对比

本文提出的算法与 LK 光流法同属于光流法，在比较匹配的运行时间时，实验控制特征点提取算法相同，均采用 ShiTomasi 算法的角点检测，利用 OpenCV 中的函数 cv::goodFeaturesToTrack，其中提取算法参数：最大角点数目 500，质量水平系数 0.01，最小距离 30，其他参数都采用 cv::goodFeaturesToTrack 默认参数。将算法循环 1000 次，取平均值，提出的算法平均运行时间为：1.361 毫秒；OpenCV 中 LK 光流法：2.892 毫秒，速度提高 52.939%，随即取 20 次实验结果如图 2 所示。

当所提出的算法和特征点法比较的时候，由于特征点法利用描述子进行匹配，所以在比较的过程中，需要加入计算描述子的时间，另外，描述子与特征点检测算法往往是一一对应的，所以最终的比较策略是将特征点提取时间加匹配时间一起进行比较。同样将算法循环 1000 次，取平均值，提出的算法平均运行时间为：8.382 毫秒；OpenCV 中 ORB 特征点法特征点提取与匹配平均时间为：35.110 毫秒，速度提高 76.125%，结果如图 3 所示。

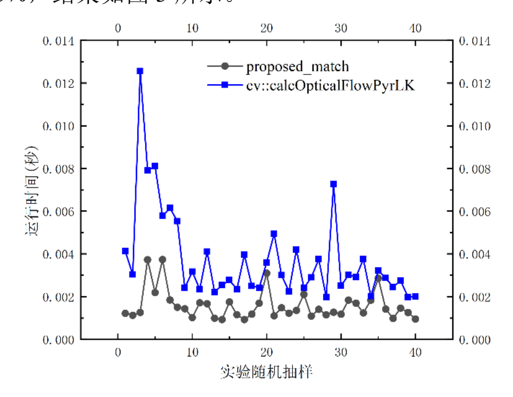


图 2　提出的算法与 LK 光流法匹配时间对比

Fig. 2　Comparison of matching time between proposed algorithm and LK optical flow method

4.2　两帧之间匹配精度评估

如果两帧的特征点 x_1 和 x_2 是正确匹配点，则应该满足对极约束：

$$x_2^T t_{21}^{\wedge} R_{21} x_1 = 0 \tag{28}$$

由于测量误差存在，上式几乎不可能成立，故求其最小二乘误差最小解：

$$all_error = \sum_{i=1}^{size} error(i)^T W^{-1} error(i) \tag{29}$$

其中 $error(i)$ 为观测数据，W 为协方差矩阵，当最小二乘解越小时，两帧之间匹配程度也越高，定义匹配分数：

$$score = \sum_{i=1}^{size} (1/(error(i)+1))^T W^{-1} (1/(error(i)+1)) \tag{30}$$

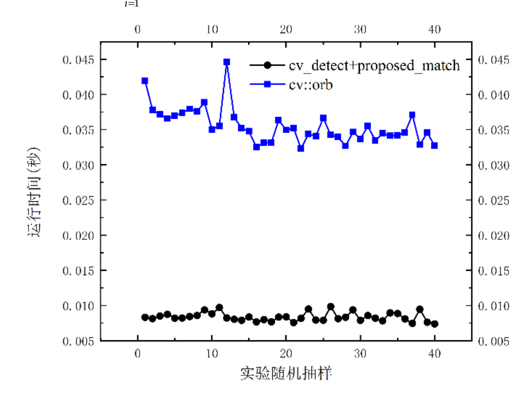


图 3　提出的算法与 ORB 特征点法检测加匹配时间对比

Fig. 3　Comparison of detection and matching time between the proposed algorithm and ORB method

根据以上定义的评估标准，计算匹配分数。如表 1 所示，本文提出的算法精度比光流法精度提高了 53.295%，这得益于提出的改进优化模型(17)，即在原有光度不变假设的基础上，引入梯度误差和噪声稀疏项。这一对比结果充分说明了改进优化模型的合理性和有效性。另外，当与 OpenCV 中 ORB 特征点检测算法相比时，虽然算法精度降低了 13.554%，但其运行时间提高了 76.125%，考虑实际情况中的实时性要求，这个精度结果也是可以被接受的。

表 1　对极约束得分

Tab. 1　Score for polar constraints

	得分	提高
proposed	28.7961	--
LK 光流	18.7847	53.295%
ORB	33.3111	-13.554%

图 4~6 分别为传统高斯牛顿法优化结果、OpenCV 中匹配优化结果以及本文提出算法匹配优化结果，展示选用的图片为 MH_02_easy 序列中相邻两张图片，并以左图特征点顺序为匹配对信息编号。图 4 在没有给定初始值的情况下，默认初始值旋转矩阵为单位矩阵，平移向量为零向量。当物体运动过快，图像模糊程度较高，旋转位姿与默认初始值偏差较大，误匹配现象较多。对比图 5、6 可以看出，由于给定了初始值进行当前帧局部邻域搜索匹配点，基本避免了特征点对的误匹配，OpenCV 中光流法仍然会出现误匹配现象，例如编号：2，11，20，21，22，27，33，39，49 等。

4.3　系统定位精度评估

本小节本文将提出的改进光流匹配算法移植到经典的 VINS-MONO 前端中，其目的是进一步验证提出的优化方法在 VI-SLAM 系统上的准确性和稳定性。为此，将改进的框架与 VINS-Mono、PL-VIO、Mono-IMU、Stereo-IMU 通过 EVO 评估工具[27]进行对比。其中，PL-VIO 是在 VINS-Mono 前端加入点线特征改进；Mono-IMU、Stereo-IMU 是 VINS-Fusion 在 VINS-Mono 基础上加入多传感器改进，并且是 KITTI Visual Odometry 榜单中排名最靠前的开源方案及改进算法。

与文献[12]一样，将对齐之后绝对轨迹误差(Absolute Trajectory Error，ATE)的均方根误差(Root Mean Square Error，RMSE)和误差均值作为评估指标。为防止结果的偶然性，进行 20 次实验取均值对比，如表 2 所示，提出的算法在 EuRoC

数据集上，均优于 PL-VIO 框架，与 VINS-Fusion 性能相近，并且经多次实验，VINS-Fusion 在 V2_03_difficult 数据集上会出现偏离真实轨迹特别远的情况。值得注意的是，由于 MH_03_medium 与 V2_01_easy 数据集在两个主导方向 x 和 y 的陀螺仪、加速度计偏置较其他数据集平均误差较大，不符合提出算法强依赖于整个系统运动的精确初始化这一基本假设，所以这两个序列性能差于 VINS-Mono。更具体而言，如果没有好的初始值，优化算法将退化为无初始化的高斯牛顿法，这不仅会增加寻优时间，也会影响系统精确度。除这两个数据集外，其他序列都有提升，最为显著的是 V2_03_difficult 序列，具有飞行器摆动较快，图像模糊程度较高等特点，是具有挑战性的数据集之一，其 RMSE 降低 18.066%，误差均值降低 28.902%。另外，图 7、8 展示了提出的算法与在 V2_03_difficult 序列估计轨迹与真实轨迹比较热力图；图 9、10 展示了 V2_03_difficult 序列中提出的算法与 VINS-Mono 算法相对轨迹误差的均方根、均值、标准差。通过这些比较，表明提出的算法在相机剧烈抖动情况下，鲁棒性更好。



图 4　无初始值的高斯牛顿迭代匹配算法

Fig. 4　Gaussian Newton iterative matching algorithm without initial values

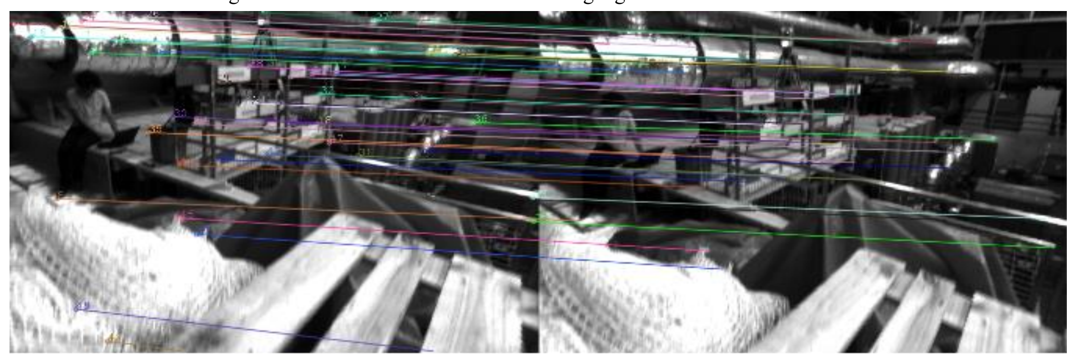


图 5　OpenCV LK 光流匹配算法

Fig. 5　Opencv LK optical flow matching algorithm



图 6　提出的匹配算法

Fig. 6　The proposed matching algorithm

表 2　多种框架性能比较

Tab. 2　Performance comparison of multiple frameworks

序列	提出的算法/(m)		VINS-Mono/(m)		PL-VIO/(m)		Mono-IMU/(m)		Stereo-IMU/(m)	
	RMSE	均值	RMSE	均值	RMSE	均值	RMSE	均值	RMSE	均值
MH_02_easy	**0.173**	**0.139**	0.194	0.164	0.186	0.158	0.184	0.152	0.184	0.151
MH_03_medium	0.408	0.322	0.402	0.314	0.422	0.334	**0.391**	**0.296**	0.468	0.416
MH_04_difficult	**0.394**	**0.325**	0.402	0.331	0.466	0.395	0.398	0.326	0.520	0.463
V1_01_easy	**0.143**	**0.119**	0.154	0.131	0.145	0.125	0.147	0.125	0.172	0.156
V1_02_medium	**0.306**	**0.249**	0.307	0.250	0.323	0.269	0.307	0.256	0.307	0.254
V1_03_difficult	0.319	0.278	0.325	0.287	0.374	0.330	0.299	0.261	**0.265**	**0.218**
V2_01_easy	0.127	0.108	**0.124**	**0.106**	0.142	0.121	**0.123**	**0.103**	0.150	0.131
V2_02_medium	**0.259**	**0.213**	0.262	0.216	0.271	0.226	0.275	0.228	0.260	0.218
V2_03_difficult	**0.322**	**0.276**	0.393	0.346	0.486	0.435	/	/	0.404	0.364

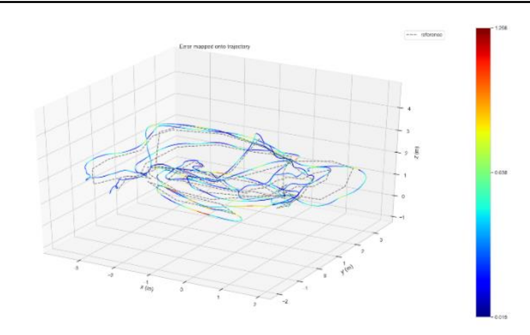


图 7　提出的算法在 V2_03_difficult 序列热力图

Fig. 7　Heat map of the proposed algorithm in V2_03_difficult sequence

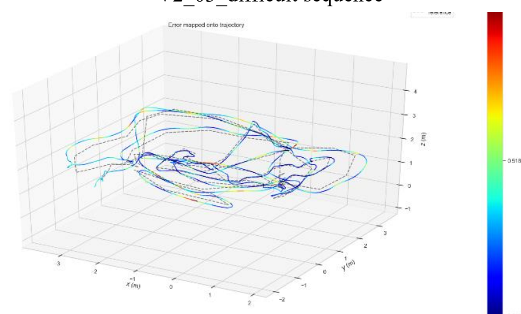


图 8　Vins-Mono 在 V2_03_difficult 序列热力图

Fig. 8　Heat map of the Vins-Mono in V2_03_difficult sequence

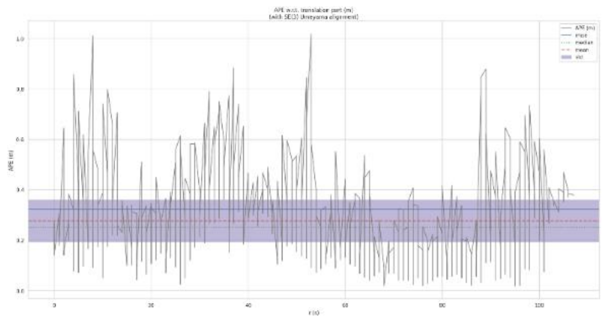


图 9　提出的算法在 V2_03_difficult 序列表现

Fig. 9　Performance of the proposed algorithm in V2_03_difficult sequence

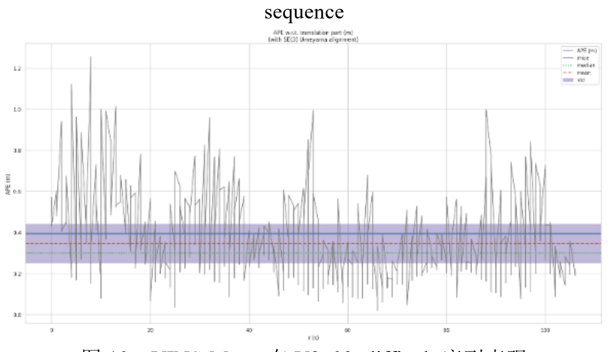


图 10　VINS-Mono 在 V2_03_difficult 序列表现

Fig. 10　Performance of the VINS-Mono in V2_03_difficult sequence

5　结束语

本文提出了一种融合视觉和 IMU 对图像进行去除运动模糊预处理，再改进光流法对特征点匹配的新算法。该算法充分利用 IMU 的运动信息，实现去除图像运动模糊信息并为迭代匹配算法提供初始值，减小特征点对误匹配概率。该算法是一个通用的模块，可以移植到其他 VI-SLAM 框架中。通过在三个不同难度层次的数据集上进行验证表明，该算法本身效率和精度都优于 OpenCV 库函数中光流匹配算法，并且将提出的算法集成到最先进的 VIO 框架 VINS-Mono 上，在位姿估计方面提高了精度。将来可以对去模糊算法进一步研究，加入正则项去除伪影，进一步提高匹配精度。

参考文献：

[1] Durrant Whyte H, Bailey T. Simultaneous localization and mapping: part I [J]. IEEE Robotics & Automation Magazine, 2006, 13 (2): 99–110.

[2] Pradeep V, Rhemann C, Izadi S, *et al.* MonoFusion: Real-time 3D reconstruction of small scenes with a single web camera [C]// Proc of IEEE International Symposium on Mixed and Augmented Reality. Australia: IEEE, 2013: 83–88.

[3] 郭金辉，陈秀万，王媛. 视觉惯性 SLAM 研究进展 [J]. 火力与指挥控制，2021，46 (1): 1-8. (Guo JinHui, Chen Xiuwan, Wang Yuan. Research progress of visual inertia slam [J]. Fire control and command control, 2021, 46 (1): 1-8.)

[4] 胡立华，左威健，张继福. 采用逆近邻与影响空间的图像特征误匹配剔除方法 [J]. 计算机辅助设计与图形学学报，2022，34 (3): 449-458. (Hu LiHua, Zuo Weijian, Zhang Jifu. The image feature mismatch elimination method of inverse nearest neighbor and influence space is adopted [J]. Journal of Computer-Aided Design and Computer Graphics, 2022, 34 (3): 449-458.)

[5] 付煜，郑爽，别桐等. 融合点线特征的视觉惯性 SLAM 算法 [J]. 计算机应用研究: 2022, 39 (2): 1-8. (Fu Yu, Zheng Shuang, Bie Tong, *et al.* Visual inertial SLAM algorithm with fused point and line features [J]. Application Research of Computers: 2022, 39 (2): 1-8.)

[6] Triggs B, Mclauchlan P F, Hartley R I, *et al.* Bundle adjustment-A Modern Synthesis [C]// Proc of International Workshop on Vision Algorithms. Heidelberg: Springer, 1999: 298–372.

[7] 伞红军，王汪林，陈久朋等. 改进误匹配剔除的 SLAM 算法 [J]. 软件导刊，2021，20 (08): 99-104. (San Hongjun, Wang Wanglin, Chen Jiupeng, *et al.* Improved SLAM algorithm for error matching elimination [J]. Software Guide, 2021, 20 (08): 99-104.)

[8] Lowe D G. Distinctive image features from scale-invariant keypoints [J]. International Journal of Computer Vision, 2004, 60 (2): 91–110.

[9] Rosten E, Drummond T. Machine learning for high-speed corner detection [C]// Proc of European conference on computer vision. Berlin: Heidelberg, 2006: 430-443.

[10] Lucas B D, Kanade T. An iterative image registration technique with an application to stereo vision [C]// Proc of Imaging Understanding Workshop, 1981.

[11] Engel J, Schöps T, Cremers D. LSD-SLAM: Large-scale direct monocular SLAM [C]// Proc of European conference on computer vision. Cham: [s. n.] , 2014: 834-849.

[12] Qin Tong, Li Peiliang, Shen Shaojie. VINS-Mono: A robust and versatile monocular visual-inertial state estimator [J]. IEEE Trans on Robotics, 2018, 34 (4): 1004–1020.

[13] Guan Qi, Wei Guoliang, Wang Licheng, *et al.* A novel feature points tracking algorithm in terms of IMU-aided Information Fusion [J]. IEEE Trans on Industrial Informatics, 2021, 17 (8): 5304–5313.

[14] Leutenegger S, Lynen S, Bosse M, *et al.* Keyframe-based visual–inertial odometry using nonlinear optimization [J]. The International Journal of Robotics Research, 2015, 34 (3): 314-334.

[15] Mourikis Anastasios I, Roumeliotis S I. A Multi-State Constraint Kalman Filter for Vision-aided Inertial Navigation [J]. IEEE International Conference on Robotics and Automation, 2007: 3565-3572.

[16] Davison A J. Active search for real-time vision [C]// Proc of Tenth IEEE International Conference on Computer Vision Volume 1. [S. l.] : IEEE, 2005: 66-73.

[17] Shi J, Tomasi C. Good features to track [C]// Proc of IEEE conference on computer vision and pattern recognition. [S. l.] : IEEE, 1994: 593-600.

[18] Fergus R, Singh B, Hertzmann A, *et al.* Removing camera shake from a

single photograph [M]// [S. l.] : ACM SIGGRAPH, 2006: 787-794.

[19] Chen Liang, Fang Faming, Wang Tingting, *et al.* Blind image deblurring with local maximum gradient prior [C]// Proc of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2019: 1742-1750.

[20] Mur-Artal R, Tardós J D. ORB-SLAM2: An open-source slam system for monocular, stereo, and RGB-D cameras [J]. IEEE Trans on Robotics, 2017, 33 (5): 1255-1262.

[21] Qin Tong, Shen Shaojie. Robust initialization of monocular visual-inertial estimation on aerial robots [C]// Proc of IEEE/RSJ International Conference on Intelligent Robots and Systems. Vancouver: IEEE, 2017: 4225-4232.

[22] Lupton T, Sukkarieh S. Visual-inertial-aided navigation for high-dynamic motion in built environments without initial conditions [J]. IEEE Trans on Robotics, 2012, 28 (1): 61–76.

[23] Shen Shaojie, Michael N, Kumar V. Tightly-coupled monocular visual-inertial fusion for autonomous flight of rotorcraft MAVs [C]// Proc of IEEE International Conference on Robotics and Automation. [S. l.] : IEEE, 2015: 5303-5310.

[24] Chen Liang, Fang Faming, Lei Shen, *et al.* Enhanced sparse model for blind deblurring [C]// Proc of European Conference on Computer Vision. [S. l.] : Springer, 2020: 631-646.

[25] He Yijia, Zhao Ji, Guo Yue, *et al.* PL-VIO: tightly-coupled monocular visual-inertial odometry using point and line features [J]. Sensors, 2018, 18 (4): 1159.

[26] Qin Tong, Pan Jie, Cao Shaozu, *et al.* A general optimization-based framework for local odometry estimation with multiple sensors [J/OL]. (2019-01-11) [2022-04-11]. http://arxiv. org/abs/1901. 03638.

[27] Grupp M. EVO: Python package for the evaluation of odometry and SLAM [J/OL]. (2017-01-01) [2022-04-11]. https://github. com/MichaelGrupp/evo.